"Stuff Goes into the Computer and Doesn't Come Out"
A Cross-tool Study of Personal Information Management

Richard Boardman
Department of Electrical and Electronic Engineering
Imperial College London
London SW7 2BT, UK
rick@ic.ac.uk

M Angela Sasse
Department of Computer Science
University College London
London WC1E 6BT, UK
a.sasse@cs.ucl.ac.uk

ABSTRACT

This paper reports a study of Personal Information Management (PIM), which advances research in two ways: (1) rather than focusing on one tool, we collected *cross-tool* data relating to file, email and web bookmark usage for each participant, and (2) we collected *longitudinal* data for a subset of the participants. We found that individuals employ a rich variety of strategies both *within* and *across* PIM tools, and we present new strategy classifications that reflect this behaviour. We discuss synergies and differences between tools that may be useful in guiding the design of tool integration. Our longitudinal data provides insight into how PIM behaviour evolves over time, and suggests how the supporting nature of PIM discourages reflection by users on their strategies. We discuss how the promotion of some reflection by tools and organizations may benefit users.

Categories & Subject Descriptors: H.5.2 User Interfaces (D.2.2, H.1.2, I.3.6) Evaluation/methodology

General Terms: Design, Human Factors, Measurement

Keywords: Personal information management, user study, files, email, web bookmarks, tool integration

INTRODUCTION

A prime characteristic of human behaviour is to acquire and keep items of value. In both the physical and digital domains, our personal spaces become populated with the objects we accumulate as our lives unfold. Personal Information Management (PIM) is an umbrella term used to describe the collection, storage, organization and retrieval of digital objects (e.g. files, addresses, and bookmarks) by an individual in their personal computing environment [11]. Bergman et al. [6] differentiate PIM from "general information management" in which a professional - such as a librarian - manages information for a range of other people with varying needs. In contrast, with PIM the onus is on the individual to manage his/her own information. Like managing one's physical possessions, PIM is frequently a burden [11,12,13], and therefore much design effort has been directed at improving PIM interfaces [5,7,8,10]. However, Whittaker et al. [14] note the relative lack of empirical research on PIM, an activity performed by *"millions of users, multiple times a day"*. This paper reports a two-phase study, aimed at improving the empirical foundation for PIM design. The study investigated the *cross-tool*, *ongoing* nature of PIM by collecting data: (1) across multiple tools (files, email and bookmarks), and (2) over time. Firstly, we outline the limitations of prior research that motivated our work.

Previous Research

Two main areas of PIM-related research can be identified: (1) empirical studies, and (2) prototype design.

Empirical Studies
Recent studies have investigated the usage of existing PIM tools, in particular email [2,13], web bookmarks [1,9] and files [3,4]. Barreau [3] presented a conceptual framework that conveys the complex, high-level nature of PIM. She outlined four component sub-activities: (1) acquisition of items to form a collection, (2) organization of items, (3) maintenance of the collection (e.g. archiving items into long-term storage), and (4) retrieval of items for reuse. Several studies have offered classifications of user behaviour in various PIM tools, focusing on the organization and maintenance sub-activities. These studies have been guided by Malone's influential work, in which he identified two fundamental strategies in office management: *filing* and *piling* [12]. Whittaker & Sidner [13] observed three email management strategies: *frequent filer*, *spring cleaner* and *no-filer*. Bälter [2] extended this classification by dividing the no-filer class into *folderless cleaner* and *folderless spring-cleaner*, depending on whether items are deleted from the inbox on a daily basis. Abrams et al. [1] described four bookmark management strategies: *no-filer*, *creation-time filer*, *end-of-session filer*, and *sporadic filer*. Barreau & Nardi [4] looked at the types of information managed by users, identifying three types based on lifetime of use: *ephemeral*, *working*, and *archived*. They noted the relative importance of ephemeral/working items retrieved by *location-based browsing*, over archived items and the use of *search*. Studies have also noted that users do not manage information simply to retrieve it later - they also store items as *reminders* of the tasks they have to perform [4,12,13].

Although previous research has made many pertinent observations and recommendations, we observe two key limitations. Firstly, findings have been fragmented along tool boundaries. Although it has been observed that people often employ multiple PIM tools in support of their high-level activities [7,10], there has been little consideration of PIM as a *cross-tool* activity. Do individuals employ similar strategies in email as in files? How are PIM tools used together? Such questions must be addressed to provide a firm empirical foundation for design work aimed at improving PIM-tool integration (see below).

Secondly, little attention has been paid to how PIM strategies change over time. One exception is Bälter [2] who proposed a model of strategy changes in email (see Figure 1). The model can be summarized in terms of two sets of strategy transitions: (1) "pro-organizing" transitions involving increases in filing tendency (solid arrows), and (2) "anti-organizing" transitions (dashed arrows). Bälter suggested that users who receive many messages might change their strategies along an "anti-organizing" path, leading to an end-state of *folderless spring-cleaner* as they file less over time. Alternatively, users might devote increased effort towards managing email and move the other way (e.g. *spring-cleaner* to *frequent-filer*). Bälter noted that further longitudinal data was needed to confirm his model. However, most work to date has been based on short-term "snapshots" of behaviour.

Prototype Design

The second area of research has focused on the exploratory prototyping of new PIM interfaces. As well as design directed at improving specific PIM tools (e.g. email), there has been extensive interest in the potential to improve integration between tools. Two main approaches can be identified in efforts to improve integration: (1) *embedding* support for managing multiple types of information within an existing tool, e.g. [5], and (2) *unifying* interaction with multiple types of information (e.g. files and email) within a consolidated interface. Examples of this second genre include *Stuff-I've-Seen* [8] which provides a unified search interface, and *UMEA* [10] which enables the organization of multiple types of information in terms of projects. Although many innovative designs have been proposed, we see a mismatch between the *tool-specific* studies that have provided observations about users' activities and problems –
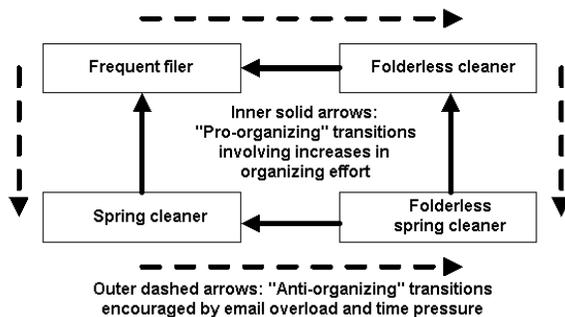

Figure 1: Model of changes in email management strategy [2]

and the substantial design effort directed at cross-tool integration. There is a need for cross-tool empirical data to provide a foundation for such cross-tool design.

STUDY OBJECTIVES AND METHOD

Our study aimed to build on previous research in two ways:

1. To provide a more effective foundation for cross-tool design, we profiled user practices across 3 commonly managed collections of personal information: files, email and web bookmarks (Phase 1, 31 participants).

2. To investigate long-term issues relating to PIM we also collected longitudinal data for 8 of the participants, again across the 3 collections (Phase 2).

Phase 1: Profiling PIM Practices

We carried out semi-structured interviews with 31 users, centered on guided tours of their file, email and bookmark collections on their main work computer. The interviews were structured to cover Barreau's four sub-activities in each tool. Since personal files are often distributed across multiple drives, in order to focus the study we asked each participant to nominate their main file collection (e.g. "My Documents", UNIX home directory, or a network drive). The use of the desktop to manage files, email or bookmarks was considered an adjunct to the respective collection. We asked participants not to tidy before the interview, which proved to be judicious (*P27: "So you know what I do now - I would have tidied it up if you'd let me"*). We performed content analysis on the interview data to extract themes relating to strategies, problems and needs. Screenshots were also captured of the desktop and the folder structures in each collection. We analyzed the folder structures to investigate: (1) the concepts used to name folders (e.g. *project*, *person*, *place*), and (2) the level of *folder overlap* (folders common to two or more collections). Finally we carried out cross-tool profiling to investigate the consistency of each participant's strategies across the collections (see "Cross-tool Profiling" section).

Phase 2 – Longitudinal Tracking of PIM Practice

Eight of the participants also took part in Phase 2, during which we tracked the evolution of the three collections and the strategies used to manage them. We developed a tool to capture snapshots of the folder structures, including counts of items within folders. Details of specific items - such as filenames - were not recorded. Participants were asked to manually initiate snapshots to lessen the infringement of their privacy. Snapshots were requested at two-week intervals over the first three months, and a final snapshot was requested five months later. Average participation was 286 days (min 218, max 309). Participants were also asked to keep a diary of significant incidents relating to the management of their files, email and bookmarks. We suggested two example incidents: creating a new folder and failing to locate an item. At the end of the study, an interview focusing on changes made to PIM strategies was carried out.

During Phase 2 we also invited the participants to try out WorkspaceMirror (WM), a software prototype developed by

the authors [7]. WM allows the user to mirror structural changes (i.e. the creation, deletion, renaming or moving of folders) between the file, email and bookmark folder hierarchies. For example, if the user creates a new file folder, he/she is asked whether equivalent email and bookmark folders should also be created. The design was motivated by observations of *folder overlap* in Phase 1 (see "Analysis of Folder Structures" section). WM was deployed as a research vehicle to explore the potential to share folder structures between PIM tools, as well as general issues related to improving integration.

Participants
All 31 participants had at least five years of computing experience, and had used their current operating system for at least one year: Windows (25), MacOS (4), and Linux (2). Eight were female, and 23 were male. The average age was 35 (min 21, max 60). 29 participants were recruited from the authors' universities. Job roles included researchers (14), students (14), and IT support (1). The remaining 2 participants were friends: one was an IT manager, and one was unemployed. Participants did not receive any incentive to take part. To lessen privacy infringements, we intentionally asked colleagues to proceed onto Phase 2. This prior familiarity established a trust basis, allowing the participants to raise concerns at any time.

RESULTS – PHASE 1
Participants were highly motivated to talk about PIM – it was an area that was important to them, and a source of problems and frustration. Despite our concerns about privacy, all participants were very open, although one (P28) joked: *"this is a high-trust exercise!"* Several participants seemed to enjoy "opening up": (*P25: "Its like a confessional getting all my computer problems off my chest"*). Only three excluded areas of their workspace for reasons of personal and/or professional confidentiality. We believe that this openness was due to participants' prior familiarity and trust in the researchers. Our Phase 1 results are presented as follows. Firstly, we compare the nature of Barreau's sub-activities [4] between the file, email and bookmark collections. We focus on organization, and offer a classification of participants' reported behaviour in each tool. Secondly, we present the results from the cross-tool profiling of participants. Finally, we summarize results relating to changes in PIM strategy.

Acquisition and Keeping
Participants stressed how the nature of acquisition varied between the tools - from manual in files and bookmarks, to uncontrolled in email: (*P11: "everything just gets stuffed into the inbox"*). All participants actively collected both files and email. File collections were highly prized, and many participants expressed the pride they felt towards the contents, much of which they had kept over a number of years: (*P29: "Some of them I'll need again, some of the things I'm quite proud of ... why should I throw it away? It doesn't cost me anything"*). Email collections were valued less than files, but many participants noted the sentimental or

professional value of a *subset* of their messages: (*P24: "I keep them to make sure I've got one thing from them to reply to. Also it's nice that the person has written"*). Bookmarks were of low importance for most participants (supporting findings in [9]), however all but one collected them to some extent. Bookmarks were valued less due to: (1) the existence of other ways of re-accessing websites, e.g. search engines, and (2) websites' ephemeral nature (*P28: "I don't trust the stability of web URLs, I would rather download the actual document"*). Bookmark collections were very small (tens of items) compared to file and email (thousands of items).

Organization
One participant (P24) succinctly summed up the ongoing challenge of PIM, and the need to organize: *"stuff goes into the computer and doesn't come out – it just builds up"*. Participants organized files most extensively, with deeper folder hierarchies, and fewer unfiled items compared to the other collections (see Table 1). We defined "unfiled" as located on the desktop or in the root folder. Since organizing strategies varied significantly between the collections, we present each in turn, starting with files.

File Management Strategies
Since no classifications of file strategies had been proposed in previous work, we developed one from scratch based on participants' strategy descriptions (see Table 2). 29 of the 31 participants had extensive folder structures, and could be divided into two groups (F1 and F2) based on the extent to which they employed a *file-on-creation* strategy (filing new items immediately). F1 users employed a predominantly *file-on-creation* strategy. They had <30 unfiled items, which tended to be left accidentally, except for a few deliberately placed work-in-progress files. F2 users filed most items on

Table 1: Extent of organization in the three collections

Phase 1 participants [n=31] unless stated otherwise	Files	Email	Bookmarks
# Folders	av. 60.4 min 5 max 235	av. 32.3 min 0 max 181	av. 17.4 min 0 max 196
Av. depth of folders	2.3	1.4	1.1
Av. # unfiled items (in root folder or on desktop)	66	828	44
Av. % collection unfiled ([n=8] Phase 2 participants)	3%	41.6%	38.8%

Table 2: File management strategies

Strategy	# users	# folders	# unfiled
F1 total filers - most items filed on creation.	18	av. 64.4 min 12 max 235	av. 14.1 min 0 max 30
F2 extensive filers – filed extensively, but left many items unfiled.	11	av. 54.8 min 28 max 124	av. 136.4 min 29 max 340
F3 occasional filers – filed very occasionally, most items unfiled, few folders.	2	av. 5 min 5 max 5	av. 240 min 150 max 330

creation but also maintained a significant subset of unfiled items (average 136.4). F2 users tended to file these items on completion of the relevant task, or during a spring-clean. Thus the location of ephemeral and working files varied between the two groups. F1 users distributed them around active folders, whilst F2 users mostly left them unfiled. However, even for the F2 users, unfiled items were a small proportion of their total file collection.

The remaining two participants (group F3) filed much less extensively, and stated that filing was not a priority. In contrast most F1/F2 users said that being organized was an important (though not always achievable) goal.

Email Management Strategies

We attempted to categorize our participants using previous strategy classifications [2,13], but were only partially successful. Therefore we developed our own classification based on participants' strategy descriptions (Table 3). Our sample included 2 *no-filers* (*folderless spring-cleaners*), and 8 *frequent filers* - but no *spring-cleaners* (users who only clean their inbox periodically). The remaining 21 participants had large inboxes (>75 items, average 1137), like the *no-filers* and *spring-cleaners* in [13], however their reported strategies did not match these classifications. They filed some new emails immediately (typically those of perceived long-term value such as e-commerce receipts), and deleted low-value spam. Other messages were left in the inbox, which was occasionally spring-cleaned. In other words they employed *multiple strategies* – a combination of frequent filer, spring cleaner, and no-filer, e.g. (*P25: "I'd like to manage as and when I receive them but I don't. I do it periodically - 10 minutes a day just to categorize the things that are important. 10 or 15 I'll categorize ... the rest of them I think oh I'll get round to doing that at some stage - but I don't normally. However I did spend an hour on a train last week tidying my emails because I was bored. I reduced my inbox by about 1500"*). We divided the 21 *multiple-strategy* participants into two sub-groups, E2 and E3, based on the extent to which they reported *manually* filing new messages on a daily basis. E2 users tried to file many emails everyday, whilst E3 users only filed a few (<5) emails. E3 users only filed specific types of message, leaving most in the inbox: (*P31: "I have a folder for registrations. I've got other [unused] folders – I don't even know what they are. The vast majority [of email] is a big long list"*). E1/E2 users indicated that organization was a priority for them, whilst E3/E4 users considered it to be less important.

Bookmark Management Strategies

We attempted to map participants' behaviour onto the classification in [1]. However, as with email, we were unsuccessful, leading us to develop a new classification (see Table 4). Only 10 participants matched a previous classification, *no filer*. The remaining 20 active collectors of bookmarks instead employed *multiple strategies*. They filed a subset of bookmarks on creation, leaving others unfiled, often as reminders, until they were spring-cleaned or simply

abandoned: (*P28: "The main thing is a mess and completely littered with things. The only exception is when I mirrored web pages for experiments. Also I keep a folder with homepages"*). We divided the *multiple-strategy* users into two groups, B1 and B2, based on the extent to which they reported filing new bookmarks on creation. Organization was of lower priority for the B2 users who had fewer folders and more unfiled bookmarks.

Analysis of Folder Structures

The previous three sections classify PIM strategies at a high level in terms of *extent of filing*. However strategy variation is not limited to extent of filing alone – it is also a question of *how* information is filed. For example a user may classify both email and files extensively, but in different kinds of folders. To explore low-level variation, we analyzed participants' folder structures to compare the concepts employed to name folders in each tool. Aggregate results are presented as follows. The most common types of file folder were *project* (short-term activities, e.g. "ucl presentation") 29%, *document class* (e.g. "letters") 17%, *role* (long-term activities, e.g. "teaching") 14%, *person* 8%, and *topic/interest* (e.g. "linux") 7%. The most common concepts for email folders were *role* 25%, *person* 19%, *project* 12%, *topic/interest* 12%, and *mailing list* 11%. For bookmarks, the most common types were *topic/interest* 55%, *document class* 13%, and *project/role* 12%.

We also investigated the extent of *folder overlap* between the collections to explore whether participants tended to create folders with the same name in different contexts. We

Table 3: Email management strategies

Strategy	# users	# folders	inbox size
E1 frequent filers - filed or deleted most messages everyday.	8	av. 50.4 min 3 max 181	av. 25.1 min 7 max 50
E2 extensive filers – aimed to file many messages everyday.	14	av. 39.9 min 8 max 91	av. 1001.6 min 87 max 5577
E3 partial filers – filed only a few (<5) messages everyday.	7	av. 5.9 min 0 max 10	av. 1367.9 min 205 max 3000
E4 no-filers - performed no active filing.	2	av. 0 min 0 max 0	av. 1105.5 min 211 max 2000

Table 4: Bookmark management strategies

Strategy	# users	# folders	# unfiled
B1 extensive filing – filed many bookmarks as they created them or at end of browsing session.	8	av. 48.4 min 6 max 196	av. 20.9 min 10 max 40
B2 partial filing – filed only sporadically.	12	av. 9.6 min 3 max 24	av. 36.9 min 7 max 100
B3 no-filers - never filed, folders abandoned. *(NB: Table does not include 1 non-collector)*	10	av. 3.3 min 0 max 26	av. 70.0 min 4 max 200

observed significant overlap for many participants, particularly between files and email. Aggregate results were as follows. The average *file/email* overlap was 7.7 folders (equivalent to 17% of file folders and 26% of email folders). The other overlaps were consistently smaller – the average *file/bookmark* overlap was 2.2 folders (7% of file folders, 23% of bookmark folders), and the average *email/bookmark* overlap was 1.7 folders (12% of email folders, 16% of bookmark folders). Most overlapping folders were named after projects (37%) or roles (32%).

Maintenance

Participants reported devoting little time to maintaining the three collections beyond occasional spring-cleans. Old items were rarely archived out of collections. Instead, it was more common for archiving to be *in situ*, e.g. several participants reported purging old messages from their inbox to a local "old-inbox" folder. Therefore all three collections tended to include a mix of ephemeral, working and archived information. Most participants reported that extensive maintenance only occurred during major life change stages such as starting a new job. Four participants reported restarting bookmark collections afresh in preference to tidying them. Our findings add to previous evidence that maintenance is of low priority [3].

Retrieval

Participants reported a strong preference for *browsing* over *search* in all three tools. This cross-tool consistency supports and extends tool-specific findings in files [4]. However there was variation between the collections in terms of the type of browsing employed. We encountered two types of browsing: (1) *location-based browsing* of folders/desktop icons [4], and (2) *sorting/scanning* of items, ordered by user-defined metadata (e.g. "name") or system-defined metadata (e.g. "size"). When retrieving files, participants employed a combination of both - browsing to a folder, and then sorting items within it. For email, retrieval was focused on sorting/scanning the inbox - location-based browsing of folders was less common. Search was used more in email than in files, but was still seen as a last resort by most participants in both collections: (*P25: "I usually know exactly where I'm going and what I'm looking for. If I search I wouldn't necessarily know the exact keyword. If you know where you're going, browsing is a lot quicker"*). Bookmark retrieval was focused on scanning recently added, or frequently accessed, items. However, many participants stated that they preferred to search the web again rather than find a bookmark they knew they had: (*P28: "If something is really exciting then I bookmark it ... when I come back to it, I just use Google"*). Nevertheless, participants continued to save bookmarks, even though many were never used. We observed similar behaviour in email, in particular the acquisition and filing of messages from mailing lists, which were never read: *(P29: "of the emails you do save, 90% you never read again"*). Similar "irrational" behaviour has been observed in paper archives, e.g. keeping personal copies of documents that are publicly available on the web [15].

In all three collections, retrieval was biased towards active and/or recently added items. However, many participants mentioned tasks that required access to older information, archived in situ, e.g. (*P22: "You look at what exam questions you had for the previous years and you decide to recycle a question or two")*. This finding is not consistent with previous claims that archived information is not useful to people [4]. We found that although older items may be accessed erratically, they can be highly valued by people (supporting findings in [15]). Interestingly, in all three collections, failure to find items appeared to happen only occasionally: (*P29: "If it exists then I'll find it. The only cases I don't is when I deleted it because I thought I didn't need it again")*. Participants indicated that mostly they *"just knew"* where to find required information. However, those rare occasions when they could not find items were highly frustrating. Three main reasons were cited for failure to find: (1) deleting items by mistake, (2) clutter, and (3) misfiling.

Cross-tool Profiling

The previous sections illustrate high-level similarities and differences in behaviour between the three collections. We also analyzed our cross-tool data on a user-by-user basis to investigate whether individual participants employed consistent strategies *across* their collections.

To do this, for each participant we produced a cross-tool profile of his/her strategies, represented by a 3-tuple (e.g. F1/E2/B1). Across the 31 participants 16 unique profiles were identified. We proceeded to cluster the profiles based on the following criterion: in which collections did the participant report making significant organizing effort? Firstly we classified each set of tool-specific strategies as either *pro-organizing* (involving high organizing effort) or *organizing-neutral* (involving low organizing effort). We attempted several classifications, from which the one shown in Table 5 emerged as the best match for the data.

Based on this classification of the tool-specific strategies, four clusters of cross-tool profiles were identified, CT1-CT4 (see Table 6). Eight participants were pro-organizing in all three tools (profile CT1), meaning that they made significant organizing effort consistently across all three collections.

Table 5: Classification of tool-specific strategies

Tool-specific strategies	Level of organizing effort
F1/F2, E1/E2, B1	"Pro-organizing", strategies that involve high organizing effort
F3, E3/E4, B2/B3	"Organizing-neutral", strategies that involve low organizing effort

Table 6: Cross-tool profiles

Cross-tool profile	# Users	% Users
CT1: pro-organizing in all 3 tools (e.g. F1/E1/B1)	8	26%
CT2: pro-organizing in files & email only	14	45%
CT3: pro-organizing in files only (e.g. F2/E3/B3)	7	23%
CT4: organizing-neutral in all tools	2	6%

The most common CT1 profile was F1/E1/B1 (four participants). Fourteen participants were pro-organizing in files and email only (CT2), with F1/E2/B2 being the most common profile (five participants). Seven were pro-organizing in files only (CT3). F2/E3/B3 and F1/E3/B2 were the most common CT3 profiles (two participants each). Two participants were organizing-neutral in all tools (CT4).

Strategy Changes: Immediate, Planned, Historical

The study had an immediate influence on the behaviour of most participants. Many indicated that taking part in the study had caused them to think more about PIM and to plan *future* strategy changes. Twelve performed ad-hoc tidying *during* the interviews, e.g. deleting files they had forgotten about. Fourteen also reported *historical* strategy changes from before the study. Five participants reported historical changes in file strategy, all of which involved increases in organization: (*P31: "I went through a phase of completely working on my desktop but it gets very cluttered"*). In email, seven participants reported historical changes - three increases and four decreases in organizing tendency: (*P12: "I used to have lots of folders for each sub-project but there wasn't enough time to manage them. Ideally there'd be rich structure but the hierarchy is now flattened"*). Five of the six reported changes in the bookmark context involved a decrease in organization (e.g. abandoning all folders).

RESULTS – PHASE 2

Eight participants proceeded onto Phase 2 of the study in which we tracked the evolution of the three collections over time. The average participation was 286 days. All eight were active collectors of files, email and bookmarks, with cross-tool profiles: CT1 (3 participants), CT2 (2 participants), CT3 (3 participants). During Phase 2 we also invited participants to try out our WorkspaceMirror prototype (WM) [7]. Four found WM useful, using it for an average of 107 days, to mirror an average of 13 newly created folders. The other four experimented with WM but did not use it in the long-term.

Growth of Collections

Table 7 summarizes collection growth over Phase 2. File collections increased in size for all eight participants, including one who archived some older material out of his file collection due to lack of space. In contrast, the other seven archived material (such as websites) *into* their file collections – contributing toward the high growth rate. Due

Table 7: Growth rates of the three collections over Phase 2

Phase 2 participants [n=8] unless stated otherwise	Av. # folders created per day	Av. # items added per day	Av. total change # folders (Phase 2)	Av. total change # items (Phase 2)
Files	0.35	5.92	100 (98.6%)	1764 (164.6%)
Email [n=4] non-archivers	0.06	5.28	12.75 (64.5%)	1551 (75.6%)
Bookmarks	0.03	0.20	9.75 (41.9%)	60.5 (41.8%)

to technical difficulties we were only able to collect email data for six participants - for whom we observed smaller folder growth compared to files. This was accompanied by a very large turnover in items. The net change in items was negative for two participants who archived an average of 3600 messages *out* of their collections, and positive for the other four participants who did not archive. Only the non-archivers are included in the email data in Table 7. Seven participants collected bookmarks very slowly (average growth: 5 folders, 25 items). One exception collected them extensively (growth: 41 folders, 306 items), but even for her, growth was much lower than in files or email.

Changes in PIM Strategy

"Non-changers"

In the closing interview, six participants said that their strategies had not changed over Phase 2. This group of "non-changers" included participants with a range of organizing tendencies. Whether pro-organizing or not, existing strategies were seen to be satisfactory or not worth changing. One observed that the effort in reorganizing his files was too high: (*P31: "Although the system I use at the moment is broken, I know how it works. To take on board a new system, I'd basically have to deal with 2 systems for a while"*). Three "non-changers" used WM, mirroring 14 new folders on average, mostly between files and email. We had anticipated that WM would stimulate pro-organizing strategy changes by allowing users to leverage filing investment in one collection across to other collections. However, these participants instead employed WM in support of existing filing strategies. Against our expectations, the two participants who made the most significant strategy changes did not consider WM to be a major contributory factor.

"Changers"

Two participants reported strategy changes. P26 (CT3: F1/E3/B2) reorganized his files two months into Phase 2 by moving all active items onto the desktop: "*I'd feel more comfortable using My Documents as an archive position and the Desktop as my working area*". Until then he had pursued a predominantly file-on-creation strategy under My Documents. He highlighted two factors that contributed to the change: (1) the need to separate active files for synchronization with his laptop, and (2) the greater influence of the interviews over our design intervention: "*Overall the tool hasn't done that much, its more the conversations between me and you. It's weird because I've become much more aware of all my directory structures*". In Phase 1 he had commented: "*I think society puts certain pressures upon people to think that being organized, being slim, being certain things are good*". This perceived social pressure might explain why most of the reorganizing happened in the context of a New Year resolution: "*I feel the need to reinvent myself, to get some good working practices together, stop drinking, stop smoking, fix bike, and organize my computer*".

P28 (CT2: F2/E2/B2) reorganized his files and email two weeks into Phase 2. In each collection he moved completed

project folders under an "old" folder: *"in the top level I have all the projects I have done in my PhD and they have become too many"*. He stated that participation in the study was a major factor in the reorganization. Although he had been planning the changes for some time, previously he had been put off by the effort involved: *"I went through the mental workload of categorizing things as important or not important* [in the study]*, so whilst this information is fresh in my memory I might as well just use it"*. He also reported an increased reliance on filing for a task which had previously been paper-based: *"the `submit here'* [folder] *is the most significant change because that is the first time that I archived stuff to remind me ... it was paper-based before"*.

For both participants, the changes, though subtle, were worthwhile: (*P28: "the [email] folders that I've created, they only take up 2% or 4% [of the inbox].* [Does that make a difference?] *Yes, because most of the stuff that comes in is day-to-day stuff I deal with today, the things I extract now are things with a longer due time"*). Both stated that the increased reflection on PIM due to the study was a key factor in the changes. Neither saw our design intervention as a major influence, although P26 did use WM to mirror 10 folders, mainly between files and bookmarks.

DISCUSSION AND CONCLUSIONS
In this section we discuss the main findings from the study and consider implications for future design.

Multiple PIM strategies
During Phase 1 we observed many participants employing multiple PIM strategies within specific collections. Previous research has also observed multiple strategies in the context of paper archives, where people tend to combine *filing* and *piling* strategies [15]. Our findings suggest that much user behaviour does not map onto earlier classifications of email and bookmark strategies [1,2,13]. Although these offer useful abstractions of PIM practice, they exaggerate the extremes – portraying users as either messy or tidy, filers or no-filers. We have attempted to classify behaviour in more detail to take account of multiple strategies.

Our cross-tool data indicates that PIM strategies also vary significantly *between* tools for many individuals. Previous work has not taken such cross-tool variation into account. Our results focused on variations in organizing strategy, e.g. participants tended to organize files more extensively than emails or bookmarks. The following factors may contribute towards such variation:

- The perceived value of information influences the selection of PIM strategies. Users feel a strong sense of ownership over files, which they have often invested significant time in authoring, and are therefore willing to take the time to organize. In contrast they feel less ownership over email, and the websites referred to by bookmarks, which are typically authored by other users.

- Organizing strategy is influenced by the likelihood and style of retrieval. Our qualitative data suggests that users are more likely to re-use files than emails or bookmarks, particularly over the long-term. Users perceive that file organization is more worthwhile since the cost of filing is offset by predicted benefits at retrieval time. Also, users tend to retrieve email by sorting on metadata, such as "sender" and "date received". Therefore there is less need to organize to facilitate folder-based browsing.

- Acquisition-related factors influence organization. Files and bookmarks are created incrementally, making them easier to organize than email, which is acquired in an uncontrolled way. Many users who would like to organize their email do not have time to do so [13].

- As well as the nature of information managed in each tool, our data suggests that a user's tendency to organize may be influenced by personality factors. Participants who stated that being tidy was important tended to be consistently pro-organizing.

Implications for Integration
Integration between PIM-tools has been repeatedly put forward as a worthy design aim [5,6,7,8,10]. Cross-tool studies can provide an empirical foundation for such design by highlighting: (1) synergies between tools that can be exploited to improve integration, and (2) differences between tool usage that may indicate barriers to integration. Our data underlines the challenge of PIM design. Designers must cater both for individual differences between users, as well as individual user's multiple strategies. Future design work must take account of the variation in strategies by providing the flexibility to manage different types of information in distinct ways. For instance, tools should give users the ability to organize information as required, whilst not penalizing those users who do not want to organize.

Our observation of folder overlap points to a subset of user activities that involve the management of multiple types of information. Most overlapping folders corresponded to *roles* and *projects*, suggesting that such categories may be usefully shared between collections, as in [10]. However, it should be emphasized that the majority of folders did *not* overlap between collections. This suggests that some production tasks are supported by single PIM tools and may not necessarily benefit from increased integration. In addition our data indicates that email contains more *contact*-based folders, whilst bookmark folders are more commonly based on user's *interests*. This variety suggests users may be constrained by designs that only support specific types of concept, such as *project* as in [10].

We also note the potential compatibility for integration of files and *filed* email. Both types of information are either self-created or assessed as having long-term value. Also folder overlap was greatest between these collections. However complete unification between files and *all* email (as pointed to by designs such as [5]) may lead to the disruption of more controlled items (e.g. files, tasks) by unprocessed email. In some cases it may be appropriate not to integrate - but to instead retain separation between tools. Planned future

work will consider integration with the other tools (e.g. calendars), and devices (e.g. PDAs) involved in PIM.

PIM as an ongoing background activity

Our study also provided insight into long-term issues regarding PIM. Although substantial historical changes were reported, including both increases and decreases in organizing tendency - the changes we observed in Phase 2 were relatively subtle pro-organizing adjustments to existing strategies. We did not observe any "global" changes in strategy along the lines of those discussed in [2], e.g. *no-filer* to *spring-cleaner*. Our experiences in Phase 2 point to the need to evaluate PIM designs over the long-term, as strategies may take a long time to evolve.

Although the observed changes were subtle, participants found them beneficial. However, the supporting nature of PIM means that users rarely devote time to planning and executing changes in strategy. Users may benefit from increased reflection with respect to PIM, so as to receive the same benefits that resulted from the "self-auditing" effect of the study. As an alternative to redesigning tools to promote reflection (e.g. providing statistics on time spent filing and searching), organizations could also play a part here. Typically, organizations are more concerned with knowledge management and other strategic IT - whilst PIM is left to the individual. Nevertheless, PIM is a key aspect of employees' activities and has the potential to cause frustration and waste time. Organizations could publicize PIM-related issues, and encourage employees to self-diagnose problems to improve their PIM effectiveness. However, managers should take care not to be overly prescriptive, or interfere with individuals' preferred style. The supporting nature of PIM leads to another dilemma for users and organizations alike: time spent thinking about PIM may result in distraction from production tasks. Tools and organizations must help the user to balance PIM and the production tasks that it supports.

The folder hierarchy is often criticized for not being easily adaptable to fast-changing user needs, and requirements for dynamic views of personal information are often emphasized in PIM design, e.g. [8]. Our findings suggest a contrasting perspective: the slow-changing nature of the hierarchy may benefit users by promoting familiarity with the personal information environment. Such familiarity in turn supports location-based finding for which users expressed a clear preference. We thus highlight persistence as an often overlooked, yet desirable design goal.

Refining the conceptual basis of PIM

We are continuing our data analysis and are looking to build on current theory in two ways. Firstly our study highlights the need for a richer descriptive vocabulary for personal information beyond technological format, and lifetime of use [4]. In particular the term "archived" is misleading, since most users do not archive explicitly. We suggest two alternative sets of terms:

1. Information usefulness: *active* (including ephemeral and working), *dormant* (inactive, potentially useful), *not useful*, and *un-assessed* (e.g. new emails).
2. Information ownership: *mine* (including self-created files, and items that have been assessed as having value, e.g. filed email), and *others* (e.g. the email inbox, and information on the internet).

Secondly, we are extending Barreau's framework [3] to reflect the cross-tool, supporting nature of PIM. Barreau conceptualized the computer as a single abstract PIM system, whereas from our data it is clear that current PIM-tools constitute a set of parallel yet inter-related systems. We also seek to modify the framework to capture the influence of production tasks in determining PIM needs.

REFERENCES

1. Abrams, D., Baecker, R. & Chignell, M. Information archiving with bookmarks: personal web space construction & organization. *Proc. CHI 1998*, 41–48.
2. Bälter, O. Strategies for organizing email messages. *Proc. HCI 1997*, Springer, 21–38.
3. Barreau, D. Context as a factor in personal information management systems. *Journal of the American Society for Information Science 46*,5 (1995), 327–339.
4. Barreau, D. & Nardi, B. A. Finding and reminding: file organization from the desktop. *ACM SIGCHI Bulletin 27*,3 (1995), 39–43.
5. Bellotti, V., Ducheneaut, N., Howard, M. & Smith, I. Taking email to task: the design & evaluation of a task management centered email tool. *Proc. CHI 2003*, 345-352.
6. Bergman, O., Beyth-Marom, R. & Nachmias, R. The user subjective approach to personal information management systems. *Journal of the American Society for Information Science 54*,9 (2003), 872-878.
7. Boardman, R., Spence, R. & Sasse, M.A. Too many hierarchies? The daily struggle for control of the workspace. *Proc. HCI International 2003*, Vol. 1, 616-620.
8. Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R. & Robbins, D. Stuff I've seen: A system for personal information retrieval & re-use. *Proc. SIGIR 2003*, 72-79.
9. Jones, W., Dumais, S. & Bruce, H. Once found, what next? A study of 'keeping' behaviors in the personal use of web information. *Proc. ASIST 2002*, 391-402.
10. Kaptelinin, V. UMEA: Translating interaction histories into project contexts. *Proc. CHI 2003*, 353-360.
11. Lansdale, M. The psychology of personal information management. *Applied Ergonomics 19*,1 (1988), 55–66.
12. Malone, T. How do people organize their desks? Implications for the design of office information systems. *ACM TOIS 1*,1 (1983), 99–112.
13. Whittaker, S. & Sidner, C. Email overload: exploring personal information management of email, *Proc. CHI 1996*, 276-283.
14. Whittaker, S., Terveen, L. & Nardi, B. Let's stop pushing the envelope & start addressing it: a reference task agenda for HCI. *Human Computer Interaction 15* (2000), 75-106.
15. Whittaker, S. & Hirschberg, J. The character, value, & management of personal paper archives. *ACM TOCHI, 8*,2 (2001), 150–17.